



OPENing the HELIX: Finding and Using Genomic Resources, a World Tour

Materials prepared by:
Warren C. Lathe III, Ph.D.
www.openhelix.com

Updated: Q3 2010





Agenda

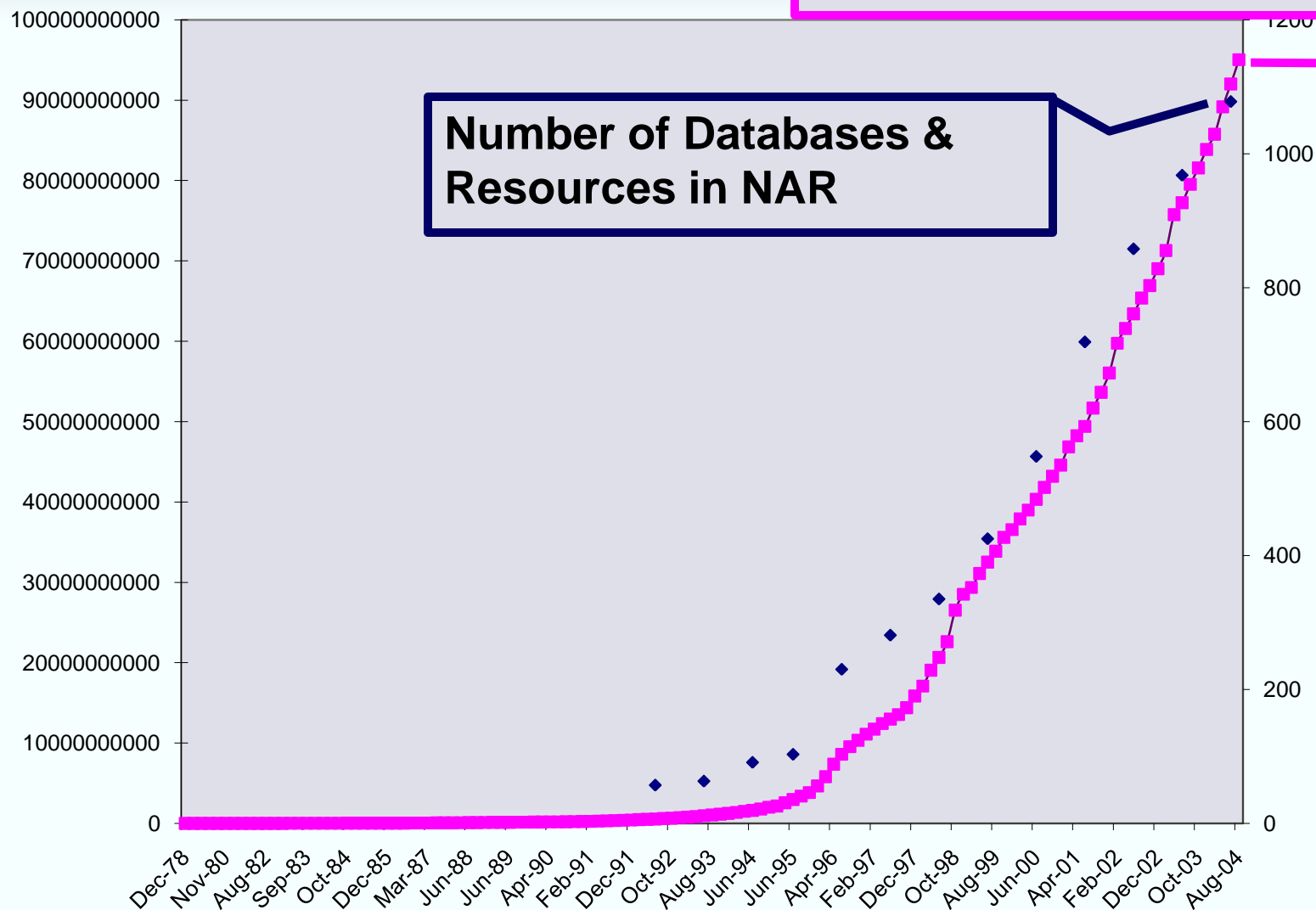
- **Introduction**
- **World Tour** of Genomics Resources
- **Finding** the Right Resource
- **Learning** How to Use the Resource
- **Conclusion & Summary**



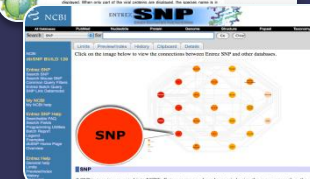
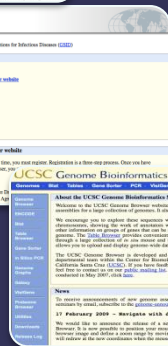
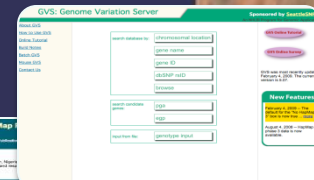
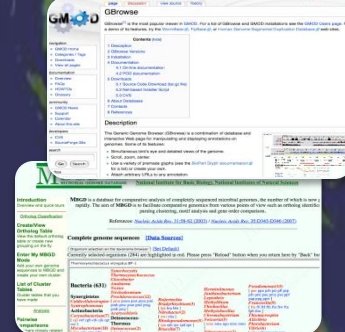
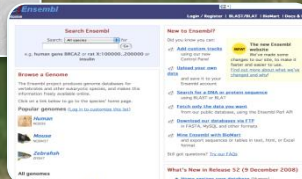
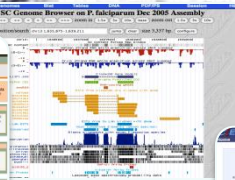
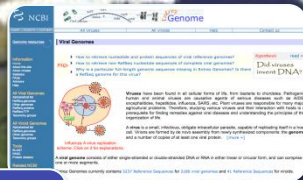
Resource growth

Sequence in data
repositories (in bps)

Number of Databases &
Resources in NAR



Worldwide Genomic Databases



Genomic Variation



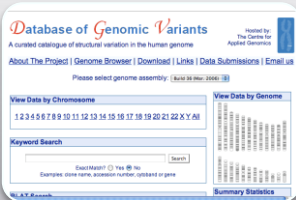
dbSNP

- <http://www.ncbi.nlm.nih.gov/SNP>



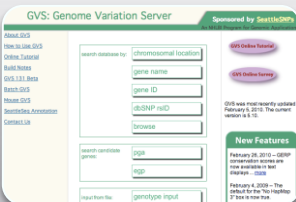
HapMap

- <http://www.hapmap.org>



DGV

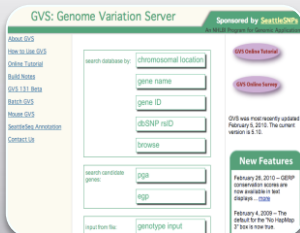
- <http://projects.tcag.ca/variation/>



GVS

- <http://gvs.gs.washington.edu/GVS/>

Genome Variation Server (GVS)



GVS

- <http://gvs.gs.washington.edu/GVS/>

GVS: Genome Variation Server

[About GVS](#)
[How to Use GVS](#)
[Online Tutorial](#)
[Build Notes](#)
[GVS 1.31 Beta](#)
[Batch GVS](#)
[Mouse GVS](#)
[SeattleSeq Annotation](#)
[Contact Us](#)

search database by: chromosomal location

gene name

gene ID

dbSNP rsID

browse

search candidate genes: pga

egp

input from file: genotype input

Select	Number Variations	Number Genotypes	Number PHASE Genotypes
<input checked="" type="checkbox"/>	169	25180	19080
<input type="checkbox"/>	166	25554	22880
<input type="checkbox"/>	162	11920	8550
<input type="checkbox"/>	162	12166	8550
<input type="checkbox"/>	104	2496	0
<input type="checkbox"/>	104	2392	0
<input type="checkbox"/>	104	2496	0
<input type="checkbox"/>	49	3479	0
<input type="checkbox"/>	48	8208	0
<input type="checkbox"/>	47	3337	0
<input type="checkbox"/>	47	3901	0
<input type="checkbox"/>	46	3542	0
<input type="checkbox"/>	45	3735	0
<input type="checkbox"/>	40	2800	0
<input type="checkbox"/>	13	296	0
<input type="checkbox"/>	13	92	0
<input type="checkbox"/>	13	85	0
<input type="checkbox"/>	13	150	0
<input type="checkbox"/>	13	85	0
<input type="checkbox"/>	7	271	0
<input type="checkbox"/>	2	24	0
<input type="checkbox"/>	1	12	0
<input type="checkbox"/>	1	12	0

default for the "No HapMap 3" box is now true.

Gene ID: 9575

Chromosome 4: 55993417 - 56107754 (-)

Allele Frequency Cutoff (%): 0 , monomorphic sites excluded , no HapMap 3 for CEU/HCB/JPT/YRI

R² Threshold for Clusters: 0.8

Minimal Genotype Coverage (%) of Snps to Be TagSnps: 85

Minimal Genotype Coverage (%) of Snps to Be Clustered: 70

Data Merging: common samples with combined variations

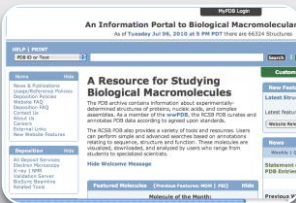
Population: HapMap-CEU(unrelated only), Submitter: CSHL-HAPMAP

Bin	Total Number of Sites	Average Minor Allele Frequency	Tag SNPs	Other SNPs
1	62	36 %	rs13132420	rs939823 [rs1464490] rs1522112 rs1522114 rs1554483 rs2035691 rs2272073 rs2412648 rs3749474 rs3805148 rs3805150 rs3805151 rs3805155 rs4340844 rs4864542 rs4864543 rs4864546 rs4864547 rs4864996 rs4864997 rs4864999 rs6843722 rs6849474 rs7658446 rs7660980 rs7677085 rs7686261 rs7691799 rs7696832 rs9312661 rs10462032 rs30817244



Pfam

- <http://pfam.sanger.ac.uk/>



PDB

- <http://www.pdb.org/>



IntAct

- <http://www.ebi.ac.uk/intact>



SMART

- <http://smart.embl.de/>

SMART

• <http://smart.embl.de/>



SMART

Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* 95, 5857-5864
 Letunic et al. (2008) *Nucleic Acids Res.* doi:10.1093/nar/gkn808

HOME SETUP FAQ ABOUT GLOSSARY WHAT'S NEW

Select your default SMART mode

You can use SMART in two different modes: **normal SMART**, the database contains Swiss-Prot, SP-TrEMBL, sequenced genomes are used; Ensembl for metazoan available [here](#).

The protein database in Normal SMART has significant domain architectures, or want to find exact domain domain annotation pages will be more accurate, and architecture query results. Remember you are exploring

Different color schemes are used to e

Normal mode

SMART MODE:

NORMAL GENOMIC

Simple

Modular

Architecture

Research

Tool

Click on the images above to select y

Information about your selected mode is stored in a [through this page](#).

You can easily change modes later, by clicking on the link in the menu:

Sequence analysis

You may use either a [Uniprot/Ensembl](#) accession number (ACC) or the protein the SMART service.

Sequence ID or ACC

TEC_HUMAN Examples: TEC_HUMAN

Sequence

Sequence SMART Reset

HMMER searches of the SMART database may also find:

- ☐ Outlier homologues and homologues
- ☐ PFAM domains
- ☐ signal peptides
- ☐ internal repeats
- ☐ intrinsic protein disorder

Domains within *Homo sapiens* protein TEC_HUMAN (P42680)

Tyrosine-protein kinase Tec

Mouse over domain / undefined region for more info; click on it to go to detailed annotation; right-click to save whole protein as PNG image

Transmembrane segments as predicted by the *TMHMM2* program (), coiled coil regions determined by the *Coils2* program (), segments of low compositional complexity determined by the *SEG* program (). Signal peptides determined by the *SignalP* program (). Intron positions are indicated with vertical lines showing the intron phase and exact position in AA.

Protein information	Domain architecture analysis
Display orthology and other data	Display all proteins with similar domain ORGANISATION or COMPOSITION. This domain architecture was probably invented with the emergence of Metazoa.
Interaction network	Pathway information
	<p>TEC_HUMAN is possibly involved in the following metabolic pathways:</p> <p>map04660: T cell receptor signaling pathway</p> <p>These assignments are based on similarity to the following orthologous groups:</p> <p>K07364: TEC (tec protein tyrosine kinase [EC:2.7.10.2])</p>

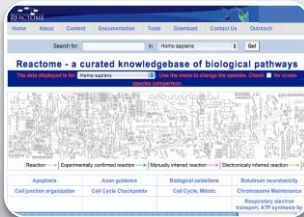
TEC_HUMAN shown as **TEC**

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the diagram. Features are also not shown when two or more occupy the same piece of sequence; the priority for display is given by SMART - PFAM - PROSITE - repeats - Signal peptide - Transmembrane - Coiled-coil - Unstructured regions - Low



KEGG

- <http://www.kegg.jp/>



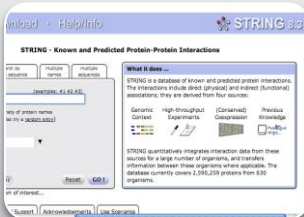
Reactome

- <http://www.reactome.org/>



MINT

- <http://mint.bio.uniroma2.it/mint/>

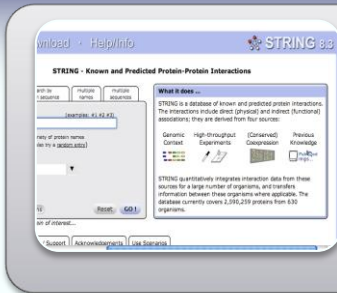


STRING

- <http://string.embl.de/>

STRING

• <http://string.embl.de/>



Home · Download ·

STRING - K

search by name search by protein sequence multi

protein name:

(STRING understands a variety of names and accessions; you can also use UniProt accession numbers)

organism:

auto-detect

interactors wanted:

COGs Proteins

please enter your protein or gene name

More Info Funding /

STRING (Search Tool for the Retrieval of Interacting Genes) is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations. They are derived from four sources:

- Genomic high-throughput experiments
- Curatorial knowledge
- Textmining
- Database cross-references

STRING quantitatively integrates interaction data from these sources for a large number of organisms, and translates information between these organisms where applicable. The database currently covers 2,090,219 proteins from 530 organisms.

Submit Annotations New Database

Your Input:

TEC Tyrosine-protein kinase Tec (EC 2.7.10.2) (631 aa) (*Homo sapiens*)

Predicted Functional Partners:

	Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	Homology	Score
DOK1									0.994
VAV1									0.989
JAK2									0.987
LYN									0.983
STAP1									0.982
KIT									0.981
PLCG1									0.945
LCP2									0.942
IL6									0.935
VAV2									0.911

DOK1 Docking protein 1 (Downstream of tyrosine kinase 1) (p62(dok)) (pp62); Docking proteins interact [...] (481 aa)

VAV1 Proto-oncogene vav; Couples tyrosine kinase signals with the activation of the Rho/Rac GTPases, [...] (845 aa)

JAK2 Tyrosine-protein kinase JAK2 (EC 2.7.10.2) (Janus kinase 2) (JAK-2); Plays a role in leptin sig [...] (1132 aa)

LYN Tyrosine-protein kinase Lyn (EC 2.7.10.2) (582 aa)

STAP1 Signal-transducing adaptor protein 1 (STAP-1) (Stem cell adaptor protein 1) (BCR downstream-sig [...] (295 aa)

KIT Mast/stem cell growth factor receptor precursor (EC 2.7.10.1) (SCFR) (Proto-oncogene tyrosine-p [...] (976 aa)

PLCG1 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma 1 (EC 3.1.4.11) (Phosphoinosit [...] (1412 aa)

LCP2 Lymphocyte cytosolic protein 2 (SH2 domain-containing leukocyte protein of 76 kDa) (SLP-76 tyro [...] (533 aa)

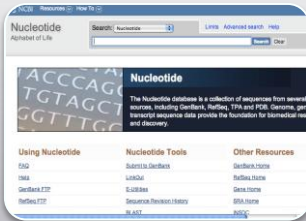
IL6 Interleukin-6 precursor (IL-6) (B-cell stimulatory factor 2) (BSF-2) (Interferon beta-2) (Hybri [...] (212 aa)

VAV2 Protein vav-2; Guanine nucleotide exchange factor for the Rho family of Ras-related GTPases (879 aa)

Views:

Neighborhood Fusion Occurrence Coexpression Experiments Database Textmining Summary Network

Nucleotides



Entrez Nucleotide

- <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide>



Entrez Gene

- <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>



Melina II

- <http://melina2.hgc.jp/>

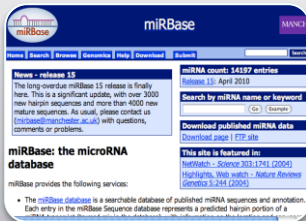



miRBase

- <http://microrna.sanger.ac.uk/>

miRBase

- <http://microrna.sanger.ac.uk/>





Home Search Browse Genomics Help Download Submit Search results

News - release 15

The long-overdue miRBase 15 release is finally here. This is a significant update, with over 3000 new hairpin sequences and more than 4000 new mature sequences. As usual, please contact us (microrna@sanger.ac.uk) with questions, comments or problems.

miRBase: the microRNA database

miRBase provides the following services:

- The [miRBase database](#) is a searchable database of published miRNA sequences and annotation. Each entry in the miRBase Sequence database represents a predicted hairpin portion of a miRNA.

Search miRBase

By miRNA identifier

Enter a miRNA accession number

Submit Reset Examine

By genomic location

Select organism, chromosome and position to retrieve all miRNAs on that region

Choose species:

For clusters

Select organism and the genomic region to search for clusters

Choose species:

By sequence

Single sequence search

Paste a sequence here

Search Results

Found 21 Homo sapiens miRNAs on chromosome 22

Stem-loop sequence MI0014246

Accession	MI0014246
ID	hsa-mir-3198
Description	Homo sapiens miR-3198 stem-loop
Stem-loop	<pre> a a c -a u a gacugugcucuc cuguuc cccag acu gcag acc g cugacacgag gguaag ggguc uga uguc ugg a a - c gg u c </pre> <p>Get sequence</p>
Coordinates (GRCh37)	22: 18246946-18247025 [-]
Overlapping transcripts	<p>sense OTTHUMT00000316178; BID-001; intron 1</p> <p>OTTHUMT00000316179; BID-002; intron 1</p> <p>OTTHUMT00000316181; BID-</p>



PubMed

- <http://www.pubmed.gov/>



iHOP

- <http://www.ihop-net.org/UniPub/iHOP/>



Textpresso

- <http://www.textpresso.org/>



XplorMed

- <http://www.ogic.ca/projects/xplormed/>

XplorMed

• <http://www.ogic.ca/projects/xplormed/>



protein tyrosine phosphatase was found to interact with mouse

XplorMed

[Home](#) [Tutorial](#) [About](#) [Contact](#)

Explore a bibliographic search in MEDLINE
step 1: Input a query in MEDLINE (a 500 abstracts)

Search MEDLINE at [GenePubMed](#) for:

Examples:

- mip AND p
- obesity AND
- ferric AND

Optionally, you can provide a session identifier (see word):

Tip: a session identifier will allow you to recover your session during one week.

Query: 278704909
step 2: Select abstracts from one or more MeSH categories

MeSH category	n	
Psychiatry and Psychology	1	19100339
Health Care	0	none
Humanities	0	none
Analytical, Diagnostic and Therapeutic Techniques and Equipment	4	19356599 19356600 19356601 19356602
Geographic Locations	0	none

After the previous selection step, your set contains now 369 abstracts.

Words ordered by association score

1.000	gene [R]
0.621	association [R]
0.466	snps [R]
0.292	polymorphism [R]
0.175	marker [R]
0.171	
0.162	
0.119	
0.117	
0.114	
0.077	
0.071	
0.067	
0.066	
0.061	
0.061	
0.060	
0.057	
0.053	

XplorMed: eXploring word context

Exploring the context of the word **haplotype**

Sentences from the abstracts containing the word.

[20145122](#)
haplotypes within linkage blocks of rxra support an approximately 30 % reduction in odds of metachronous neoplasia arising in the proximal colon among carriers of specific haplotypes , which was strongest orproximal , 0.67 ; 95 % ci , 0.52 0.86 for carriers of a cggcca **haplotype** .

[20131239](#)
RESULTS : two **haplotype** blocks that correspond approximately to fam167a and blk were observed .

[20064547](#)
however , one **haplotype** allele of nfe2l2 was associated with 2 years earlier age at AD onset pc=0.013 and 4 years earlier age at surgery for posterior subcapsular cataract pc=0.019 . another **haplotype** of nfe2l2 was associated with 4 years later age at surgery for cortical cataract pc=0.009 .

[19767079](#)
Alpha: the nlrp3 **haplotype** of the 2 snps also showed a significant association .

Score: 0.05 (0.0-0.9)

Control modify Alpha: Score:

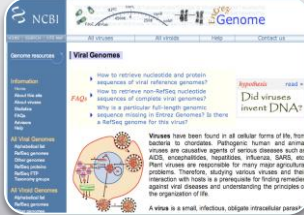
NEXT ACTION: [Compute chains of related words](#) [Reset](#)

Save selected abstracts
Output format is
readable
Tip for EndNote users
Explore context of any word

OHRI Bioinformatics



Genome Databases (prokaryotes & viruses)



NCBI Viral Resources

- <http://www.ncbi.nlm.nih.gov/genomes/GenomesHome.cgi?taxid=10239>



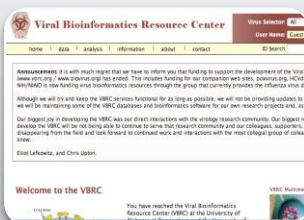
CMR

- <http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi>



IMG

- <http://img.jgi.doe.gov/cgi-bin/pub/main.cgi>

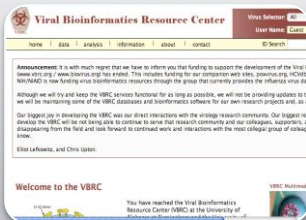


VBRC

- <http://www.vbrc.org/>

VBRC

• <http://www.vbrc.org/>



Announcement: It is with much regret that we have to inform you that funding to support the development of the Viral Bioinformatics Resource Center (VBRC) at the University of Illinois at Chicago (UIC) is now ending. The NIH/NIAID is now funding virus bioinformatics resources through the group that currently provides the influenza virus database.

Although we will try and keep the VBRC services functional for as long as possible, we will not be providing updates to it. We will be maintaining some of the VBRC databases and bioinformatics software for our own research projects and, as a result, we will be unable to respond to requests for assistance.

Our biggest joy in developing the VBRC was our direct interactions with the virology research community. Our biggest regret is that the VBRC will be no longer able to continue to serve the research community and our colleagues. Support is disappearing from the field and looking forward to continued work and interactions with the most collegial group of colleagues.

Elliot Lefkowitz, and Chris Upton

Viral Bioinformatics RESOURCE CENTER

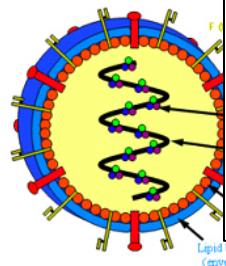
[Virology.ca Tools](#) » [Organisms](#) »

GATU

Please read: NEWSFLASH.

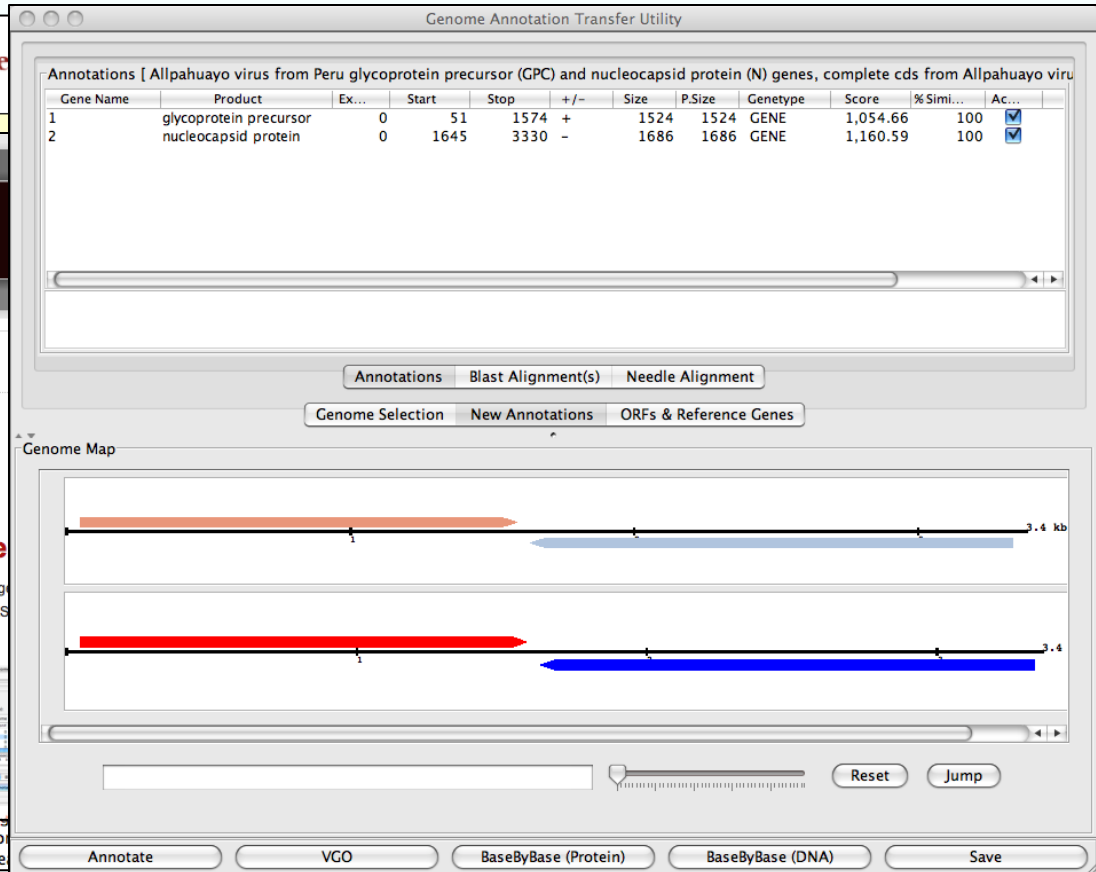
Genome Annotation Transfer

Genome Annotation Transfer Utility (GATU) annotates a genome by transferring annotations from a reference genome. The reference genome proteins/mature peptides of the reference genome are BLASTed against the target genome to identify genes/mature peptides in the genome to be annotated.



Structure of hepatitis viruses. Hepatitis virus is a type species.

possible bioterrorist threats or re-emerging infectious diseases.



Genome Annotation Transfer Utility

Annotations [Allpahuayo virus from Peru glycoprotein precursor (GPC) and nucleocapsid protein (N) genes, complete cds from Allpahuayo virus]

Gene Name	Product	Ex...	Start	Stop	+/-	Size	P.Size	Genotype	Score	% Simi...	Ac...
1	glycoprotein precursor	0	51	1574	+	1524	1524	GENE	1,054.66	100	<input checked="" type="checkbox"/>
2	nucleocapsid protein	0	1645	3330	-	1686	1686	GENE	1,160.59	100	<input checked="" type="checkbox"/>

Annotations Blast Alignment(s) Needle Alignment

Genome Selection New Annotations ORFs & Reference Genes

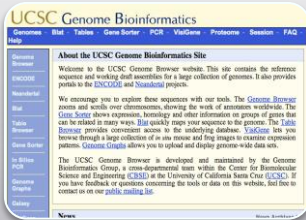
Genome Map

4 kb

Reset Jump

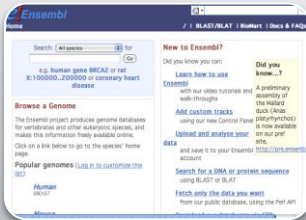
Annotate VGO BaseByBase (Protein) BaseByBase (DNA) Save

Genome Databases (eukaryotes)



UCSC Genome Browser

- <http://genome.ucsc.edu>



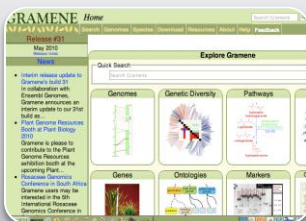
Ensembl

- <http://www.ensembl.org>



Gbrowse

- <http://gmod.org/wiki/Gbrowse>

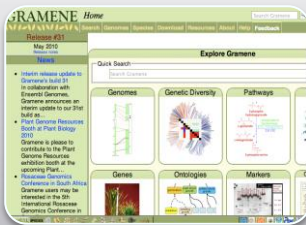


Gramene

- <http://www.gramene.org/>

Gramene

• <http://www.gramene.org/>



GRAMENE

Release #31
May 2010
Release notes

News

- Interim release update to Gramene's build 31
In collaboration with Ensembl Genomes, Gramene announces an interim update to our 31st build as...
- Plant Genome Resources Booth at Plant Biology 2010
Gramene is please to contribute to the Plant Genome Resources exhibition booth at the upcoming Plant...
- Rosaceae Genomics Conference in South Africa
Gramene users may be interested in the 5th International Rosaceae Genomics Conference in Cape Town...

Subscribe to RSS Feed

Have Questions...?

- Quick Search Help
- Ask questions through [Feedback](#) or [Email](#).
- See [FAQ](#).

GRAMENE Genes

Search Genomes Species Download Resources About Help Feedback

Gene Home | Search | Submit | Help | Tutorial | FAQ | Release Notes

Browse Ontology Database by: [Trait](#) | [Plant Structure](#) | [Growth Stage](#)

Find: In: Type: Species: With Phenotype ☐

E.g., [d1](#), [dwarf-1](#), [GR:0060184](#), [flower*](#), [*tillering stage*](#) Or [view help](#)

Genes 1 to 25 of 171. of 7. | [Next](#)
[\[Download Data \]](#)

Species	Gene Symbol	Gene Name	Synonyms	TO	PO	GRO	GO	EO	Chr. No.	Gene Type
Oryza sativa	ADL1	ADAXIALIZED LEAF 1	adl1, adaxialized leaf-1	leaf shape, stem length	2 formation of axillary shoot, stem, leaf production, inflorescence emergence from flag leaf sheath, LP.06 six leaves visible, booting, 5 fruit formation, stem elongation, leaf, 6 ripening, LP.05 five leaves visible	04-stem elongation stage, 05-booting stage, 03-tillering stage, 02-seedling, 06-heading stage, 08-dough stage, 07-milk stage	multicellular organismal development			Not sequenced

Home Page: [Simple Search](#)

Help

[Help](#)

[Tutorial](#)

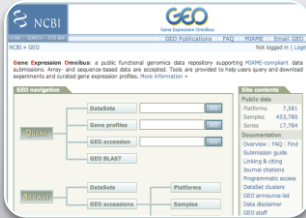
[FAQ](#)

[Release Notes](#)

Submit Genes

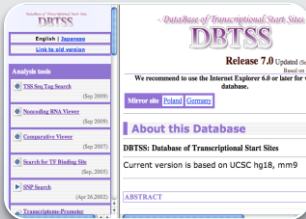
[Submit a gene](#)

Expression



GEO

- <http://www.ncbi.nlm.nih.gov/geo>



DBTSS

- <http://dbtss.hgc.jp/>



ArrayExpress

- <http://www.ebi.ac.uk/arrayexpress>




DAVID

- <http://david.abcc.ncifcrf.gov/>

DAVID

• <http://david.abcc.ncifcrf.gov/>





Home Start Analysis Short

Upload List Background

Gene List Manager

Select to limit annotations by one or more species [Help](#)

- Use All Species -
Homo sapiens(398)
Unknown(5)

Select Species

List Manager [Help](#)

demolist1
demolist2

Select List to:

Use Rename
Remove Combine

Show Gene List

[View Unmapped Ids](#)

Functional Annotation Clustering

Current Gene List: demolist2
Current Background: Homo sapiens
379 DAVID IDs

☒ **Options** **Classification Stringency** Medium

[Rerun using options](#) [Create Sublist](#)

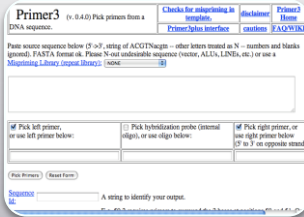
178 Cluster(s) [Download File](#)

Annotation Cluster 1	Enrichment Score: 4.2		Count	P_Value	Benjamini
<input type="checkbox"/> GOTERM_BP_FAT	response to wounding	RT	34	9.2E-7	2.2E-3
<input type="checkbox"/> GOTERM_BP_FAT	inflammatory response	RT	23	1.8E-5	6.4E-3
<input type="checkbox"/> GOTERM_BP_FAT	defense response	RT	32	1.2E-4	2.5E-2
<input type="checkbox"/> GOTERM_BP_FAT	immune response	RT	33	4.3E-4	4.5E-2
<input type="checkbox"/> KEGG_PATHWAY	Cytokine-cytokine receptor interaction	RT	20	1.1E-3	1.4E-1

Annotation Cluster 2	Enrichment Score: 3.84		Count	P_Value	Benjamini
<input type="checkbox"/> GOTERM_BP_FAT	regulation of apoptosis	RT	42	7.3E-6	8.9E-3
<input type="checkbox"/> GOTERM_BP_FAT	regulation of programmed cell death	RT	42	9.2E-6	5.6E-3
<input type="checkbox"/> GOTERM_BP_FAT	regulation of cell death	RT	42	9.9E-6	4.8E-3
<input type="checkbox"/> GOTERM_BP_FAT	anti-apoptosis	RT	17	5.1E-5	1.5E-2
<input type="checkbox"/> GOTERM_BP_FAT	positive regulation of programmed cell death	RT	26	7.1E-5	1.9E-2
<input type="checkbox"/> GOTERM_BP_FAT	positive regulation of cell death	RT	26	7.6E-5	1.9E-2



Analysis and Algorithms



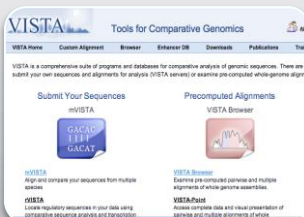
Primer3

- <http://frodo.wi.mit.edu/>



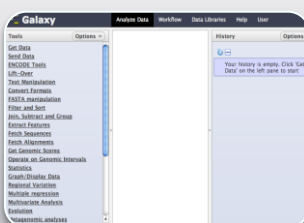
ClustalX

- <http://bips.u-strasbg.fr/fr/Documentation/ClustalX/>



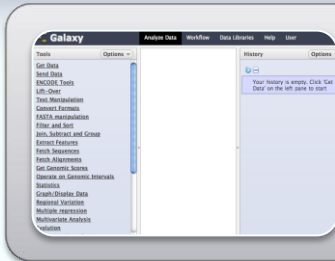
VISTA

- <http://genome.lbl.gov/vista/index.shtml>



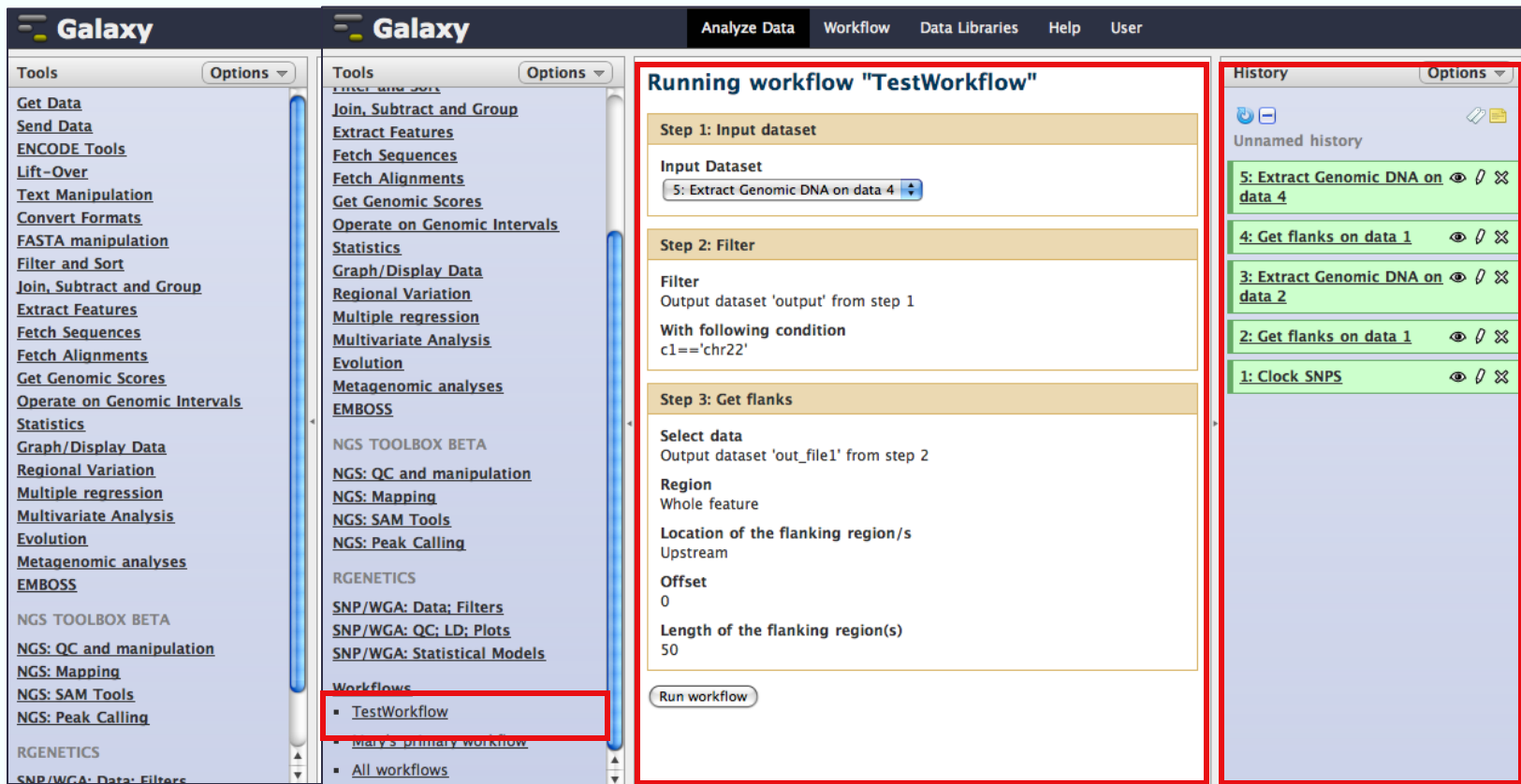
Galaxy

- <http://www.usegalaxy.org/>



Galaxy

- <http://www.usegalaxy.org/>



Galaxy Analyze Data Workflow Data Libraries Help User

Tools Options

- Get Data
- Send Data
- ENCODE Tools
- Lift-Over
- Text Manipulation
- Convert Formats
- FASTA manipulation
- Filter and Sort
- Join, Subtract and Group
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Metagenomic analyses
- EMBOSS
- NGS TOOLBOX BETA
- NGS: QC and manipulation
- NGS: Mapping
- NGS: SAM Tools
- NGS: Peak Calling
- RGENETICS
- SNP/WGA: Data; Filters
- SNP/WGA: QC; LD; Plots
- SNP/WGA: Statistical Models
- Workflows
- TestWorkflow**
- Mary's primary workflow
- All workflows

Running workflow "TestWorkflow"

Step 1: Input dataset

Input Dataset
5: Extract Genomic DNA on data 4

Step 2: Filter

Filter
Output dataset 'output' from step 1
With following condition
c1=='chr22'

Step 3: Get flanks

Select data
Output dataset 'out_file1' from step 2
Region
Whole feature
Location of the flanking region/s
Upstream
Offset
0
Length of the flanking region(s)
50

Run workflow

History Options

Unnamed history

- 5: Extract Genomic DNA on data 4
- 4: Get flanks on data 1
- 3: Extract Genomic DNA on data 2
- 2: Get flanks on data 1
- 1: Clock SNPS



Agenda

- Introduction
- **World Tour** of Genomics Resources
- **Finding the Right Resource**
- **Learning** How to Use the Resource
- Conclusion & Summary

Search for Resources: www.openhelix.com

The screenshot shows the OpenHelix website search interface. A red box highlights the top navigation bar, including the OpenHelix logo, a 'Login | Register' link, and a 'Menu' section with links to 'Search', 'Recent Queries', 'Tutorial Suites', 'Catalog', 'By Category', and 'Free Tutorials'. The main search area is titled 'Find and Use the Genomics Resources Relevant to Your Needs' and contains a search input field, 'Search' and 'Clear' buttons, and links to 'Advanced search options' and 'Help'. A large blue arrow points to the 'Search' button, and a speech bubble contains the text 'tag AND snps'. Below the search bar, a red box highlights the 'Advanced search options' link. The left sidebar contains 'Latest Blog Post' (Friday SNPets, What's Your Problem? Open Thread) and 'OpenHelix Highlights' (New Tutorial added: PSI, SGKB, Ensembl Legacy, GenMAPP; New Galaxy/OpenHelix publication in Current Protocols: Galaxy: A). The right sidebar contains a 'Help' section titled 'Searching help and hints:' with a list of search tips. At the bottom, a red box highlights the 'Advanced search options' link, and a blue box highlights the 'Help' link. A white box with a red border contains the 'Results per page:' dropdown (set to 10) and the 'Sort results based on:' dropdown (set to 'Combined relevance'). The footer contains the copyright notice '© 2010 OpenHelix, LLC. All rights Reserved.' and the text 'design & development: biobyte solutions'.

OpenHelix

Login | Register

Menu

Search
Recent Queries
Tutorial Suites
Catalog
By Category
Free Tutorials

Find and Use the Genomics Resources Relevant to Your Needs

Enter words, phrases, or sentences that describe what you are looking for, we'll perform a relevancy search on our tutorial suites and hundreds of bioinformatics resources to find the ones most relevant to your needs. Or try these sample queries: SNPs, GBrowse, interactions, NCBI, "copy number variation".

Search Clear

Advanced search options Help

Find and Use the Genomics Resources Relevant to Your Needs

Enter words, phrases, or sentences that describe what you are looking for, we'll perform a relevancy search on our tutorial suites and hundreds of bioinformatics resources to find the ones most relevant to your needs. Or try these sample queries: SNPs, GBrowse, interactions, NCBI, "copy number variation".

Search

Advanced search options Help

Latest Blog Post

- Friday SNPets
- What's Your Problem? Open Thread

OpenHelix Highlights

- New Tutorial added: PSI, SGKB, Ensembl Legacy, GenMAPP
- New Galaxy/OpenHelix publication in Current Protocols: Galaxy: A

Help

Searching help and hints:

- Simple logic operators can be used to connect the search terms:
 - AND:** both terms must be present (for example: *human AND genome*). This is the default mode.
 - OR:** either of the terms must be present (for example: *human OR sapiens*)
 - AND NOT:** second term must not be present (for example: *vertebrata AND NOT mammalia*)
- Look for phrases using double quotes (for example: "gene expression")
- Use the advanced options to adjust the number of results shown and change the sorting options

Advanced search options Help

Results per page: 10

Sort results based on:

- ✓ Combined relevance
- Resource web pages relevance
- OpenHelix training material relevance
- Resource PageRank

Expression Omnibus tutorials

- Updated Online Tutorial for GeneTests

© 2010 OpenHelix, LLC. All rights Reserved. design & development: biobyte solutions

Find Resources: www.openhelix.com

OpenHelix.

Menu

Search

Recent Queries

Tutorial Suites

Catalog

By Category

Free Tutorials

Subscription Information

Other Training/Materials

Quick Reference Cards

Curriculums

Newsletter

Seminars/Workshops

Site Documentation

OpenHelix Genomics Blog

About OpenHelix

Contact and Feedback

Latest Blog Post

- Friday SNPs
- What's Your Problem? Open Thread

OpenHelix Highlights

- New Tutorial added: [PSI, SGKB, Ensemble Legacy, GenMAPP](#)
- New Galaxy/OpenHelix publication in Current Protocols: [Galaxy: A Web-Based Genome Analysis Tool for Experimentalists](#)

Latest News

- Updated Online Tutorials for NCBI resources including an NCBI Overview and PubMed and the Gene Expression Omnibus tutorials
- Updated Online Tutorial for GeneTests

tag AND snps

Search

Clear

Advanced search options

Help

Your query 'tag AND snps' matched 1793 documents in 41 resources.

Free OpenHelix tutorials that match your search

Links open in: ☒ current window ☐ new window

Page 1 of 5

OpenHelix Tutorials	Resource Information
	NIEHS SNPs Program: http://egp.gs.washington.edu/ NIEHS (National Institute of Environmental Health Sciences) SNPs Program: resources and databases to determine the relationship between genetic variability and response to environmental exposures <ul style="list-style-type: none">OpenHelix tutorial materials relevant to your searchWeb pages from the resource that were relevant to your search
	GVS: http://gvs.gs.washington.edu/GVS/ GVS, Genome Variation Server: tool to access and analyze human genotype data from dbSNP and HapMap <ul style="list-style-type: none">OpenHelix tutorial materials relevant to your searchOpenHelix Blog posts relevant to your searchWeb pages from the resource that were relevant to your search
	Tagger: http://www.broad.mit.edu/mpg/tagger/ A tool to select tag SNPs (Single Nucleotide Polymorphisms) for genome-wide association studies (GWAS) <ul style="list-style-type: none">Web pages from the resource that were relevant to your search
	HapMap: http://hapmap.org/ A multi-country consortium project to identify and catalog human genetic variation and catalog the SNPs <ul style="list-style-type: none">OpenHelix tutorial materials relevant to your searchWeb pages from the resource that were relevant to your search
	GeneSNPs: http://www.genome.utah.edu/genesnps/ GeneSNPs (Single Nucleotide Polymorphisms) database catalogs human gene sequence and polymorphism data, focusing on genes important in the mediation of environmental exposure <ul style="list-style-type: none">OpenHelix tutorial materials relevant to your searchWeb pages from the resource that were relevant to your search
	CGAP: http://cgap.nci.nih.gov/ The National Cancer Institute's (NCI) Cancer Genome Anatomy Project, CGAP: datasets and analytical tools for the study of gene expression in many types of cancer including the SAGE Anatomic Viewer (SAV) for visual display of gene expression in human normal and malignant tissues <ul style="list-style-type: none">OpenHelix tutorial materials relevant to your search
	SeattleSNPs: http://pga.gs.washington.edu/ SeattleSNPs (Single Nucleotide Polymorphisms): focused on genetic variability in genes important to the inflammatory pathway in humans <ul style="list-style-type: none">OpenHelix tutorial materials relevant to your searchWeb pages from the resource that were relevant to your search
	PlasmoDB: http://plasmodb.org/ PlasmoDB, Plasmodium genome database <ul style="list-style-type: none">Web pages from the resource that were relevant to your search
	ApiDB: http://apidb.org/ ApiDB / EuPathDB, eukaryotic pathogens database resource for genomic-scale datasets associated with Cryptosporidium, Giardia, Leishmania, Plasmodium, Toxoplasma, Trichomonas and Trypanosoma <ul style="list-style-type: none">Web pages from the resource that were relevant to your search
	HMDB: http://www.hmdb.ca/ HMDB, Human Metabolome Database for small molecule metabolites in the human body <ul style="list-style-type: none">Web pages from the resource that were relevant to your search

Page 1 of 5

© 2010 OpenHelix, LLC. All rights Reserved.

design & development: babybyte solutions

Find Resources: www.openhelix.com

Resources
that match
query

OpenHelix Tutorials

Resource Information



NIEHS SNPs Program: <http://egg.gs.washington.edu/>

NIEHS (National Institute of Environmental Health Sciences) SNPs Program: resources and databases to determine the relationship between genetic variability and response to environmental exposures

- ☐ OpenHelix tutorial materials relevant to your search
- ☐ Web pages from the resource that were relevant to your search



GVS: <http://gvs.gs.washington.edu/GVS/>

GVS, Genome Variation Server: tool to access and analyze human genotype data from dbSNP and HapMap

- ☐ OpenHelix tutorial materials relevant to your search
- ☐ OpenHelix Blog posts relevant to your search
- ☐ Web pages from the resource that were relevant to your search

Tagger: <http://www.broad.mit.edu/mpg/tagger/>

A tool to select tag SNPs (Single Nucleotide Polymorphisms) for genome-wide association studies (GWAS)

- ☐ Web pages from the resource that were relevant to your search



HapMap: <http://hapmap.org/>

A multi-country consortium project to identify and catalog human genetic variation and catalog the SNPs

- ☐ OpenHelix tutorial materials relevant to your search
- ☐ Web pages from the resource that were relevant to your search



GeneSNPs: <http://www.genome.utah.edu/genesnps/>

GeneSNPs (Single Nucleotide Polymorphisms) database catalogs human gene sequence and polymorphism data, focusing on genes important in the mediation of environmental exposure

- ☐ OpenHelix tutorial materials relevant to your search
- ☐ Web pages from the resource that were relevant to your search



CGAP: <http://cgap.nci.nih.gov/>

The National Cancer Institute's (NCI) Cancer Genome Anatomy Project, CGAP: datasets and analytical tools for the study of gene expression in many types of cancer including the SAGE Anatomic Viewer (SAV) for visual display of gene expression in human normal and malignant tissues

- ☐ OpenHelix tutorial materials relevant to your search

Available
tutorials



Find Resources: www.openhelix.com

GVS: <http://gvs.gs.washington.edu/GVS/>

Go to resource

GVS, Genome Variation Server: tool to access and analyze human genotype data from dbSNP and HapMap

OpenHelix tutorial materials relevant to your search

Tutorial movie (audio-visual presentation)

Slide 27: When determining **tag SNPs**, GVS uses an algorithm that is applicable only to one population.

Slide 30: ... stringent you wish the potential **tag SNPs** to be. The r-squared threshold gives a minimal...

Slide 32: ... displays include genotypes, **tag SNPs**, linkage disequilibrium and a SNP summary.

Slide 34: ... **SNPs** in the bin that are **tag SNPs**. The same color-code applies here as it does in the previous...

Slide 38: ... a quick example looking for **tag SNPs** in two populations. This will allow us to learn more...

Slide 39: ... a complete and combined set of **tag SNPs** that would cover variation in both populations to use...

Slide 40: ... in the Allele Frequency Cutoff box. To obtain the data, step 4, we click display **tag snps**.

Slide 41: ... open a new window with the **tag SNPs** that we can use to cover the variation in both populations.

Slide 42: ... down further, you will find the **tag SNPs** and bins for each of the two populations you view...

Slide 44: ... bins highlighted by bars above. The **SNPs** that can be used for **tag SNPs** are shown with...

Slide 45: ... and clicking the display **tag snps** button, we will get the warning message shown here...

5 additional slides match the query

[View the tutorial movie](#)

[Go to tutorial info page](#)

View movie

Training Materials

Query match
in OH training
materials

Training materials (access animated slides, handouts, exercises)

OpenHelix Blog posts relevant to your search

Post title
[Personal Genomics, clinical
assessment and online resources](#)
... information and does a quick **tag SNP**
analysis using GVS. That tutorial is only one
very small...

Tags
SIFT Polyphen OMIM personalized medicine UniProt
NIEHS SNPs Lancet gvs PharmaGKB HGMD
genomics pubmed

Query match
in OH Blog

Web pages from the resource that were relevant to your search

Title	Size
FAQ	21.4 kB
The MultiPop-TagSelect Algorithm	13.6 kB
Genome Variation Server	20.1 kB
How to Use GVS	36.4 kB
Genome Variation Server	31.1 kB

Showing only top 5 hits. [Display all 6 hits](#). OpenHelix is not responsible for the content of external internet sites.

Query match
on resource
website



Agenda

- Introduction
- **World Tour** of Genomics Resources
- **Finding** the Right Resource
- **Learning** How to Use the Resource
- Conclusion & Summary



Training Materials: Tutorial

GVS

Tutorial and training materials by OpenHelix

Introduction to: Learn to use **GVS** (Genome Variation Server), a comprehensive database that provides quick access to all data found in both the dbSNP and HapMap databases and includes specialized data from other projects. The database includes a suite of analysis tools such as linkage disequilibrium plots, tag SNPs, merging populations, and more. In addition you can upload our

Genome Variation Server (GVS)
Genome Variation Server (GVS)

Online Tutorial S

You'll learn:

- to search for v
- to obtain haplo
- to download s

More about the r

Genome Variation Se
model the associatio
Genomic Applications

Launch

Online Tutoria

Related tutor

This tutorial is a

SeattleS
NIEHS S

Categories

View additional

Variation

GVS

[Introduction](#)

Search

Populations

Parameters

Analysis Results

Example: Merging Populations

Summary

Exercises

Download Materials

GVS: Genome Variation Server

GVS: Genome Variation Server

Sponsored by **SeattleSNPs**
An NHLBI Program for Genomic Applications

SeattleSNPs
Variation Discovery Resource

Materials prepared by:
Warren C. Lathe, PhD
www.openhelix.com

Tutorial

Version 2

OpenHelix

00:06/03:57

Training Materials: Slides

GVS
Tutorial and training materials by OpenHelix

Introduction to:

Genome Variation Server (GVS)

Online Tutorial Suite

You'll learn:

- to search for variations
- to obtain haplotype
- to download specific

More about the resource

Genome Variation Server (GVS) is a web-based model the associations between genetic variation and Genomic Applications (PGA)

Launch Online Tutorial

Related tutorials

This tutorial is a part of

SeattleSNPs
NIEHS SNPs

Categories

View additional tutorials

Variation & Analysis

Learn to use **GVS** (Genome Variation Server), a comprehensive database that provides quick access to all data found in both the dbSNP and HapMap databases and includes specialized data from other projects. The database includes a suite of analysis tools such as linkage disequilibrium plots, tag SNPs, merging populations, and more. In addition you can upload our own data and use the GVS analysis and visualization tools. If your research involves genetic variation and SNPs, learn how to

gvs_slides_v2.ppt

Slides

Open Table

Since we have chosen two populations and have merged the data by combining samples, we have invoked the "MultiPop algorithm." As mentioned earlier in this tutorial, when combining samples, GVS performs the tag-SNP selection for each group individually. It then uses the resulting tag SNPs in the MultiPop-TagSelect algorithm to determine tag SNPs for the combined groups.

A new window will appear with three links. The first link will be tables of tag SNPs calculated by the MultiPop algorithm. The next two links are graphic representations of tag SNPs from the two different populations you chose.

We will click the first "open table display of tag SNPs" link.

Slide 40 of 56



Training Materials: Exercises



The Genomics Knowledge You Need, When You Need It

www.openhelix.com

Step-by-Step checklist/instructions for the GVS introduction exercises

1.) Find the number of individuals in the HapMap-JPT population that are homozygote for the rare allele (at a minimum frequency of 20%) of one nonsynonymous SNP located in the coding region of the ABO gene. What are their genotypes?

Step	Action	✓
1	Go to the GVS homepage: http://gvs.gs.washington.edu/GVS	
2	Click on the "Gene Name" button in the "search database by" section.	
3	In the resulting window, type in "abo" in the text box labeled "gene name" (leaving the "Beyond Your Target" boxes empty) and click "Search."	
4	Check the HapMap-JPT (Japanese) population data source in the "Select Population" table. Note: Make sure to uncheck other checked populations. The "PGA-EUROPEAN-PANEL" population will be checked by default and checking another population will not uncheck it.	
5	In the "Set up parameters for display and analysis" table, type in "20" for the "Allele Frequency Cutoff (%)" box (in the "Filtering SNPs" row). Change the "Output SNPs By" menu to "Chromosome Position." Leave all other parameter choices as default.	
6	Click the "display genotypes" button (green) at bottom of the page.	
7	In the resulting browser window, click the "open graphical display of genotypes" link. Scroll down to bottom of the graphic display window and look at color legend. The color for homozygote-rare allele is yellow. The color for SNP function in a coding-nonsynonymous region is red.	
8	Scroll back up to view the graphic. There are 5 SNP(s) listed here that are coding nonsynonymous based on the color-code. We'll look at the first listed. The first coding nonsynonymous SNP number is at location 135121293.	
9	Viewing the column of the SNP(s) we've chosen, there are three individuals that have the box color for the homozygote-rare allele. The individual designations are 1411:5204, 1411:5219 and 1411:5236.	
10	Bring the text/table data display window (by clicking the "open table display of genotypes" link from step 7) to your view and look for the SNP(s) listed that you found in step 8 (hint: the same color-code for SNP function applies here as it does in the graphic display). Find the individual(s) you determined in step 9. Their genotype is T/T.	

ve database that provides quick access to all data found in both data from other projects. The database includes a suite of merging populations, and more. In addition you can upload our your research involves genetic variation and SNPs, learn how to

omic Applications (PGA), which works to identify, genotype and tional Heart Lung and Blood Institute's (NHLBI) Programs for

ownload

on Exercises V
(PDF file)

Exercises

he other tutorials in the group interesting:

nvironmental Genome Project (EGP) SNPs


Subscribed Tutorials: www.openhelix.com/subscriptions.shtml

- Tutorials and Training Materials on nearly 100 resources including:






















- dbSNP
- NCBI Entrez
- dbGaP
- STRING
- Ensembl
- Genetics Home Reference
- GIBBS Motif Sampler
- miRBase
- PFAM & many more...

List of all OpenHelix tutorials

This is the full list of tutorials offered by OpenHelix, grouped by category. Please note that a tutorial can appear in multiple categories.

Click on the  icons to expand/collapse individual categories. Clicking on a tutorial name will display detailed information, together with a list of all available training materials.

Expand all
















Name	Description	Resources
 Algorithms and Analysis		
 Expression		
 Allen Mouse Brain Atlas	Mapped gene expression data in mouse brain	Allen Mouse Brain Atlas
 Alternative Splicing and Transcript Diversity (ASTD) database	A bioinformatics resource for alternative splice events and transcripts for human, mouse, and rat	ASTD
 ArrayExpress	A public repository for microarray gene expression data at the EBI	ArrayExpress
 CleanEx	A Database of Heterogeneous Gene Expression Data Based on A Consistent Gene Nomenclature	CleanEx
 DAVID	A tool that analyzes large lists of genes to provide biological meaning	DAVID
 DBTSS	Database of Transcriptional Start Sites	DBTSS
 Gene Expression Omnibus (GEO)	A gene expression/molecular abundance repository and a curated, online resource for gene expression data	Gene Expression Omnibus (GEO)
 GENSAT	Provides an extensive amount of high quality images of gene expression in the central nervous system of the mouse.	GENSAT
 PhenomicDB	Phenotypes database	PhenomicDB
 UCSC Genome Browser: The Additional Tools	Additional tools at the UCSC Genome Browser	UCSC Genome Browser
 VISTA	Tools for Comparative Genomics	VISTA
 Genome Databases (eu)		
 Genome Databases (pro)		
 Literature and Text Mining		
 Miscellaneous		
 Nucleotides		
 Pathways		
 Proteins		
 Variation & Medical		

Sponsored Tutorials:

www.openhelix.com/cgi/freeTutorials.cgi

■ Sponsored Resource Tutorials

- Galaxy
- Vista
- PDB
- SGKB
- GBrowse
- SGD
- FlyBase
- WormBase
- Rat Genome Database
- Mouse Genomics Informatics
- ZFIN

Freely available OpenHelix tutorials					
This is the full list of freely available tutorials offered by OpenHelix. Clicking on a tutorial name will display detailed information, together with a list of all available training materials.					
Name	Description	Category	Resources	Sponsored By	
 FlyBase	A resource for the genes, genome and molecular biology of <i>Drosophila melanogaster</i> and related species.	Genome Databases (eu)	FlyBase	-	
 Galaxy	Analysis tools for researchers	Algorithms and Analysis	Galaxy	-	
 GBrowse	GBrowse User Introductory Tutorial	Genome Databases (eu), Genome Databases (pro)	GBrowse	-	
 Gramene	A resource on rice and other grass genomes	Genome Databases (eu)	Gramene	Gramene	
 Mouse Genome Informatics (MGI)	The Mouse Genome Informatics resource provides data, tools, and analyses for the mouse model organism.	Genome Databases (eu)	Mouse Genome Informatics (MGI)	-	
 NIEHS SNPs	National Institute for Environmental Health Sciences Environmental Genome Project (EGP) SNPs	Variation & Medical	NIEHS SNPs Program	-	
 Overview of Genome Browsers	Various Genome Browsers examined	Genome Databases (eu), Genome Databases (pro)	GBrowse IMG Ensembl UCSC Genome Browser NCBI Map Viewer	UCSC Bioinformatics Group	
 PSI SGKB	The Protein Structure Initiative Structural Genomics Knowledgebase	Proteins	The Protein Structure Initiative Structural Genomics Knowledgebase	PSI SGKB	
 Rat Genome Database (RGD)	Rat Genome Database	Genome Databases (eu)	Rat Genome Database (RGD)	-	
 RCSB PDB	RCSB Protein Data Bank	Proteins	RCSB PDB	RCSB PDB	
 Saccharomyces Genome Database (SGD)	Saccharomyces Genome Database	Genome Databases (eu)	SGD	-	
 UCSC Genome Browser: An Introduction	The UCSC Genome Browser Introduction	Genome Databases (eu), Algorithms and Analysis	UCSC Genome Browser	UCSC Bioinformatics Group	
 UCSC Genome Browser: Custom Tracks and Table Browser	UCSC Genome Browser advanced topics	Genome Databases (eu), Algorithms and Analysis	UCSC Genome Browser	UCSC Bioinformatics Group	
 UCSC Genome Browser: The Additional Tools	Additional tools at the UCSC Genome Browser	Expression, Genome Databases (eu), Nucleotides, Proteins, Algorithms and Analysis	UCSC Genome Browser	UCSC Bioinformatics Group	
 VISTA	Tools for Comparative Genomics	Expression, Nucleotides, Variation & Medical, Algorithms and Analysis	VISTA	-	

OpenHelix Blog: blog.openhelix.com

- Daily news and information about genomics data, research and resources
- Weekly video tips on how to use database and tool features
- Weekly “What’s Your Problem” thread to ask questions on the use of resources and where to find them
- More...

The screenshot displays the OpenHelix Blog homepage. At the top, the title "The OpenHelix Blog at OpenHelix" is prominent. A navigation bar lists various resources: bioinformatics, chv, comparative genomics, conferences, databases, disease, ENCODE, ensembl, evolution, galaxy, GBrowse, GenBank, genetics, genome, genomes, genomics, genomics resources, GINA, gramine, GWAS, hapmap, img, literature, MGI, mouse, NCBI, nhgri, nih, OHIM, personal genomics, plants, proteins, pubmed, Reactome, research, RGD, snps, training. A search bar is located on the right. The main content area features a "What's your problem? Open Thread" dated 13 May, 2010, by Mary. Below it is a "Tip of the Week: Chromhome, for karyotype level comparative genomics" dated 12 May, 2010, also by Mary. A "Share This" button and a "Comments" section are visible. On the right sidebar, there are sections for "OpenHelix News", "OH Blog Feeds", "Brought to you by:", "Research Blogging", "About", "Latest Tweet", and "Latest Tips". A small inset image shows a "Chromosome Homology Mapping Atlas" with a table of data and a small portrait of a person.



Agenda

- Introduction
- **World Tour** of Genomics Resources
- **Finding** the Right Resource
- **Learning** How to Use the Resource
- **Conclusion & Summary**

Worldwide Genomic Databases

